Amendments to the Specification:

Please amend the title of the application as it appears on page 1 of the originally filed application as follows:

SYSTEM FOR CREATING A <u>SPACE-EFFICIENT</u> DOCUMENT CATEGORIZER <u>FOR TRAINING AND TESTING OF AUTOMATIC CATEGORIZATION ENGINES</u>

Please add the following new paragraphs at page 2 after paragraph [0004] and before the BRIEF DESCRIPTION OF THE DRAWINGS of the originally-filed application.

SUMMARY OF THE INVENTION

Embodiments of the invention overcome one or more deficiencies in the prior art by providing, among other things, a space-efficient database system optimized for the process of categorizing pages and creating a stable training and testing environment for the generation of automatic categorization engines. The present invention provides a database that is stable and independent of any changes that may occur to the contents of an original page. Accordingly, a method (or an apparatus) for providing a space-efficient database system optimized for the process of categorizing pages is desired to address one or more of these and other disadvantages.

In accordance with one aspect of the invention, a method comprises receiving a locator of a network resource and determining if a database already contains stored information derived from the network resource at a previous point in time, effectively freezing the network resource to the previous point in time. The method further comprises storing information derived from the network resource pointed to by the locator of the network resource upon determination that the database does not contain stored information derived from the network resource at the previous point in time. The process of storing

2

comprises creating a copy of at least a portion of the network resource pointed to by the locator and writing the copy to the database.

In accordance with another aspect of the invention, an apparatus comprises a storage medium having stored therein a plurality of programming instructions and a processor coupled to the storage medium to execute the programming instructions. The programming instructions are designed to (i) receive a locator, (ii) determine if information is previously stored; and (iii) store information generally as set forth above.

Alternatively, the invention may comprise various other methods and apparatuses.

Other features will be in part apparent and in part pointed out hereinafter.